International Journal of Semantic Computing
Vol. 4, No. 3 (2010) 301–330

DOI: 10.1142/S1793351X10001036

World Scientific
www.worldscientific.com

SEMANTIC INTEGRATION IN GEOSCIENCES

ZAKI MALIK

Department of Computer Science
Wayne State University, Detroit, MI 48202, USA
zaki@wayne.edu
http://www.cs.wayne.edu/∼zaki

ABDELMOUNAAM REZGUI

School of Information Sciences
University of Pittsburgh, Pittsburgh, PA 15260, USA
arezgui@sis.pitt.edu

BRAHIM MEDJAHED

Department of Computer and Information Science
University of Michigan, Dearborn, MI 48128, USA
brahim@umich.edu

MOURAD OUZZANI

Cyber Center, Discovery Park
Purdue University, West Lafayette, IN 47907, USA
mourad@cs.purdue.edu

A. KRISHNA SINHA

Department of Geosciences
Virginia Tech, Blacksburg, VA 24060, USA
pitlab@vt.edu

We present an approach for the semantic integration of geoscience data, and a system implementing this approach. Specifically, we demonstrate the use of data ontologies and application of markup languages for semantic integration of data and services. We introduce a domain level object ontology, called Earth and Planetary ONTology (EPONT) to explore, extract, and integrate information from heterogeneous geologic data sets. As proof of concept, we define the DIA engine, an extensible infrastructure for the Discovery, Integration, and Analysis of geoscience data, tools, and services. DIA provides a collaborative environment where scientists can share their resources (e.g., geochemical data, filtering services, etc.) by registering them through well-defined ontologies. We envision the DIA infrastructure to also use other classes of ontologies, namely process and service, for knowledge creation.

Keywords: Geoinformatics; semantic integration; ontology; markup.

1. Introduction

In recent years, geoscientists have moved towards using the Web as a medium to exchange and discover vast amounts of data. The current practice is dominated by establishing methods to access data with little emphasis on capturing the meaning of data to facilitate interoperability and integration. Some common methods for integration include schema integration leading to the use of mediated schemas that provide a uniform query interface for multiple resources [31]. Methods using peer data management [3, 77] allow participating data sources to retrieve data directly from each other, holding the promise to extend data integration to the Internet scale. However, such query capabilities require syntactic and semantic mapping across resources to be effective. Multi-level *ontologies* are therefore a pre-requisite for semantic integration. In this paper, we adopt the definition of an ontology as a set of knowledge terms, including the vocabulary, the semantic interconnections, and rules of inference and logic for some particular topic [29]. This paper emphasizes the need to develop community-based transformative capabilities that would result in a semantic interoperability and integration infrastructure for data and knowledge sharing. We use the geosciences domain to illustrate key findings and presented approach.

It is well-recognized that semantic integration enables scientists to advertise, query and access massive amounts of heterogeneous data on an unprecedented scale [41, 49]. To facilitate semantic integration, we propose an ontology-based interoperability approach whereby individual scientific communities collaboratively develop multiple levels and classes of ontologies. Multiple levels of ontologies facilitate the presentation of information to humans as well as machines leading to automated query capabilities within the emerging *Semantic Web* vision [7]. Since a large proportion of data on the Web nowadays is mostly understandable by humans or custom developed applications, it makes the integration process inefficient, and error-prone. Thus, it is difficult for other scientists to correctly understand the semantics of the data, leading to poor use of the Web in answering complex science queries. For instance, "how can a geoscientist get all the data that link volcanism to climate change?" Hence adding semantics to enable machines to understand and automatically process the data that they merely display at present is essential [49].

Semantics-aware techniques and concepts provide an abstraction layer above existing IT technologies (e.g., databases, applications, etc.), which bridges the connection of data, tools (content), and various processes across scientific domains and IT silos. From a human perspective, added semantics enable relevant and intelligent interactions than those available with the traditional point-to-point integration approaches [33]. Other advantages of adopting the Semantic Web technologies would include: facilitated knowledge management (processes of capturing, extracting, processing and storing knowledge) [18], integration across heterogeneous domains (e.g., using ontologies) [12, 17], composition of complex systems [54, 68], ability to handle non-textual items as images, multimedia, etc. [32, 64], efficient information filtering

(sending selective data to right clients), machine understanding (ability to take humans out of the "integration loop"), forming of virtual communities (e.g., geoscientists using specific ontologies) [61], serendipity (finding unexpected collaborators), and vocabulary standardization. In light of the above discussion, we suggest that there is an immediate need to define models, encodings, and techniques that facilitate semantic and syntactic interoperability among geoscientists.

The IUGS (International Union of Geological Sciences) Commission for the management and application of Geoscience Information (CGI) has taken a step towards data interoperability and integration related to geological maps among geoscientists. CGI has formed an "interoperability working group" to establish an initial data model and XML based exchange language. The resulting language is referred to as the GeoScience Mark-up Language (GeoSciML). Several similar initiatives (XML-based languages) for hydrology, images, and chemicals [50] emphasize the importance of developing a common knowledge interface for information understanding and exchange. We propose that similar techniques for all geoscience disciplines (e.g., petrology, geophysics, mineralogy, etc.) be adopted to facilitate intra- and inter-discipline semantic interoperability. For instance, languages such as StructML (for geological structures), RockMinML (for rocks and minerals) would need to be defined and used to create a common understanding across the broad geoscience disciplines. However, in the face of well-developed sub-disciplines within the geosciences, complex geoscience queries require that apart from intra-discipline understanding, inter-discipline integration be also carried out. For example, consider the geosciences query "*What is the distribution of U/Pb zircon ages of A-Type plutons in Virginia. Identify the correlation between these plutons and their geophysical* (*e.g.*, *gravity*) *properties*". This query requires inter-disciplinary integration between three geoscience disciplines, namely, geochemical, geophysical, and geochronological data sets (for identifying A-Type plutons, showing gravity contours, and specifying zircon ages, respectively). Moreover, for visual analysis, the results need to be shown in the form of a *Virginia* geologic map and a 3-D model. As mentioned earlier, various geoscientists in a particular discipline may have collected data related to the above query. Intra-discipline integration would be facilitated if the *A-Type filter tool* (that differentiates among A, I, S, and M-type plutons) can operate on the different geochemical datasets in a "standard" manner, without code modification for different data formats, conventions, and meanings. In essence, such integration requires semantics-based search and information brokering, which is facilitated through inter-ontology relationships (ontologies defined for each discipline) [37].

We identify three types of ontologic frameworks for discovery and integration: Object (e.g. materials), process (e.g., chemical reactions), and service (e.g., simulation models) ontologies [45, 73]. Objects represent our understanding of the state of the system when the data were acquired, while processes capture the physical and chemical forcings on objects that may lead to changes in state and condition over time. Service provides systematic approaches using tools (simulation models and

analysis algorithms) to assess multiple hypotheses, including inference or prediction. These three classes of ontologies within the semantic layer of the scientific cyberinfrastructure are required to enable automated discovery, analysis, utilization, and understanding of data through both induction and deduction, advancing computational thinking along the pathway from data to knowledge, and ultimately to insight. Although this paper emphasizes object ontology and its relationship to markup languages in the geosciences for integrative purposes, we recognize the need to expand this capability where scientists can examine the relationships between data and external factors such as processes that may influence our understanding of "why" certain events happen. We emphasize the need to go from analysis of data to concepts and data inherent in thermodynamics, kinetics, heat flow, mass transfer, and other similar areas of interest. The development of object ontologies is a pre-requisite for semantic interoperability across process, object and service ontologies.

Object ontologies can be represented at three levels of abstraction: upper, mid, and domain-level ontologies [66]. Upper-level ontologies [57] are domain independent and provide universal concepts applicable to multiple domains. Mid and domain-level ontologies represent concepts that are both domain specific and are linked to higher level ontologies for semantic integration across multiple resources. Earlier, we introduced an Earth and Planetary ONTology (denoted EPONT) as a domain-level ontology for efficient, reliable, and accurate data sharing among geoscientists [72, 74]. EPONT utilizes existing community-accepted high level ontologies such as SUO[a] (Semantic Upper Ontology: IEEE endorsed), SWEET[b] (Semantic Web for Earth and Environmental Terminology) and NADM[c] (North American Geological Data Model). In particular, the SWEET ontology contains formal definitions for terms used in earth and space sciences and encodes a structure that recognizes the spatial distribution of earth environments (earth realm) and the interfaces between different realms. These earth realms have associated properties with appropriate units and provide an extensible mid-level terminology. EPONT supports extension of these concepts to domain-specifics to which data are registered.

In this paper, we describe our approach of using domain ontologies as contained in "mark-up languages initiative" (similar to GeoSciML, HydroXC, etc.) towards more expressive capabilities through the use of deeper ontologies. Specifically, we describe the DIA engine: a semantically-enabled system for the Discovery, Integration, and Analysis of geoscience data. DIA uses EPONT for registering and discovering geoscience data. DIA is a service-based system for answering complex inter-domain geoscience queries. It uses various geoscience tools encoded as Web services (both in-house and off-site) to overcome syntactic and semantic heterogeneity problems.

[a] http://suo.ieee.org/
[b] http://sweet.jpl.nasa.gov/ontology/units.owl
[c] http://pubs.usgs.gov/of/2004/1334

The rest of this paper is organized as follows. In Sec. 2, we provide an overview of markup languages and ontologies. We show how these technologies can be used in complement to facilitate interoperability and integration in the geosciences. In Sec. 3, we describe the DIA system. We describe its architecture and show how it is used in practice. Section 4 presents an overview of related work, and Sec. 5 concludes the paper.

2. Advancing Geoscience Research through Markup Languages and Ontologies

The extent, complexity, and sometimes primitive form of existing data sets and applications, as well as the need for the optimization of the collection of new data dictate that only a well-coordinated and sustained effort through the emerging science of geoinformatics will allow the community to attain its scientific goals. Such an effort requires that integration be performed at both the intra-discipline and inter-discipline levels. We propose to map markup languages to appropriate levels of ontologies, where ontologies can be instrumental in inter-discipline integration through semantic description of the data. Since it is unreasonable to assume a single ontology (or markup language) to cover all sub-fields of geoscience, integration between different disciplines can be supported by defining inter-ontology relationships and mappings, and computing the integrated concept hierarchy [10].

Ontologies allow the use of formal and descriptive logic statements which permits more expressive query capabilities for data integration through reasoning. An ontology reasoner (e.g., *Pellet*[d]) is a service that takes statements encoded in an ontology as input and infers new statements from them. In their support for reasoners, Horrocks *et al.* state that "understanding is closely related to reasoning" [35]. Other uses of reasoners include (but not limited to): revealing relationships (superclass/subclass) among classes, determining the most specific types of individuals, detecting inconsistent class definitions, etc. [17]. Projects like the Suggested Upper Merged Ontology (SUMO) [52] and the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [46] provide a framework for developing mid-level and domain ontologies, thus promoting data interoperability, automated inference, and natural language processing. Following similar lines, as an initial step towards realizing the goal of taking geoinformatics forward to the Semantic Web, we have defined EPONT to aid the markup languages initiative.

Figure 1 shows the high-level representation of the planetary ontology. For example, the package "planetary material" can be used to represent the nature (physical, chemical) of substances and their properties. This figure also shows the utilization of imported and inherited properties from additional packages, e.g., Physical Properties, Planetary Location, and Planetary Structure, to more fully define the concept of Planetary Materials. Existing ontologies available from SWEET ontology library

[d] http://pellet.owldl.com/

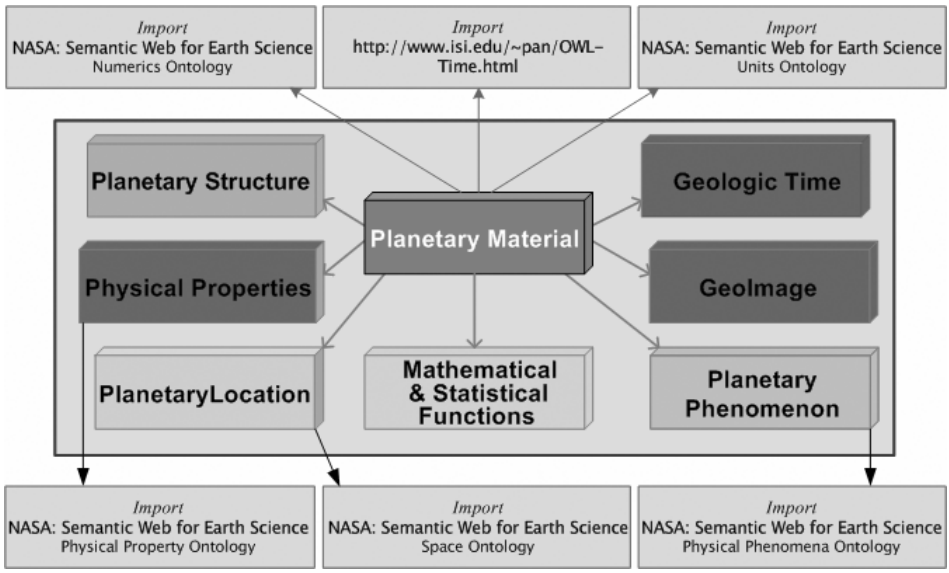

Fig. 1. Earth and Planetary Ontology (EPONT) packages.

such as Numerics, Time, and Units are also used as they provide common concepts which are useful for the development of data level ontologies.

2.1. *Intra-discipline integration using markup languages*

Markup languages facilitate interoperability by making the data processable in both human and machine-readable formats [16, 32, 54, 55, 61, 67, 68]. XML is the most widely used markup language, designed to facilitate the sharing of data across different information systems, particularly those connected via the Internet [11]. Geoscientists will need to share a common data transfer model in various disciplines, so that intra-discipline data exchange becomes easier. Note that the idea of developing markup languages for the geoscience disciplines is not completely new and several efforts in this regard are well under way. For example, scientists have developed markup languages for hydrology (e.g. HydroXC), images (e.g. IIML), chemicals [50], and geological interpretations [70]. The advantages of developing XML-based markup languages include: the ability to represent the most general computer science data structures as records, lists and trees, a simple document structure, strong syntax and parsing requirements which make the associated algorithms simple and efficient, the hierarchical structure of the documents which is suitable for most types of documents [62] . In the following, we elaborate on the importance of markup languages with the help of GeoSciML: a markup language primarily developed for sharing geological interpretations.

With interoperability as its foremost objective, the GeoSciML initiative aims to develop a conceptual model of geological maps that borrows from existing data

models and ontologies [70]. GeoSciML uses the domain modeling methodology described in ISO 19109 in particular, and is based around the notion that an application schema defines the "feature-types" for a domain. The General Feature Model (GFM) is GeoSciML's underlying meta-model, which is expressed in UML. GeoSciML uses the UML to XML conversion rules from Geography Markup Language (GML) for the canonical XML serialization, but other serializations are also possible. Future development for other MLs could employ the same methodology as GeoSciML, i.e., use the GFM and UML, and then generate a GML-based serialization to develop the MLs. This would make the integration of these easier in an inter-disciplinary system using a standard mapping method. However, as mentioned earlier, ontologies should be used to define individual data items so that the associated meaning can be understood in and across different domains / disciplines. Thus, one can think of MLs as standardized interfaces, linked to more expressive ontologies using first-order/descriptive logic [5, 6] as reasoning tools to dig deep for individual data items for semantic discovery and integration. UML may act as the unifying language for linking MLs and ontologies. The UML class diagrams provide a common visual presentation of ML structures (e.g., schema) and ontology concepts, thus enabling the linkage of semantics and implementation in a unified tool environment [17].

Figure 2 shows the role of markup languages and ontologies in integrating heterogeneous data. Since both the markup language and the ontology will have similar terms (but at different levels of granularity), data translators or mapping artifacts need to be written so that the intended data items can be found. In essence, this means that data made accessible over the Web through MLs needs to be semantically annotated, using formal ontologies if we are to take geosciences to the Semantic Web. Other scientific applications focusing on data integration as well as peer-to-peer data management systems have already benefited from such techniques [31]. Semantic annotation and mapping is a time consuming process which needs to consider a variety of aspects that are useful in understanding the semantics of a schema including data values, element names, data/relationship constraints, structural information, domain knowledge, and cardinality relationships. Therefore, automatic methods are required for schema matching. A number of such approaches have been proposed in the literature, ranging from schema matching using different types of evidences to identify mappings [59, 69] to focusing on models represented in a specific modeling language [19, 22, 34, 36, 40, 44]. Similarly, [3, 39] present approaches for such mappings that essentially connect paths in the ML to chains of properties in an ontology.

Figure 3 shows a subset of the UML model for GeoSciML (the class diagram for "earth materials" is shown). Similarly, Fig. 4 shows the corresponding EPONT ontology developed as a result of a "concentrated discipline-based community effort" with a detailed level of abstraction. GeoSciML provides similar abstraction details but at a different level, not shown in Fig. 3. This facilitates the mapping process between markup languages and ontologies. For example, *grossChemistry* of

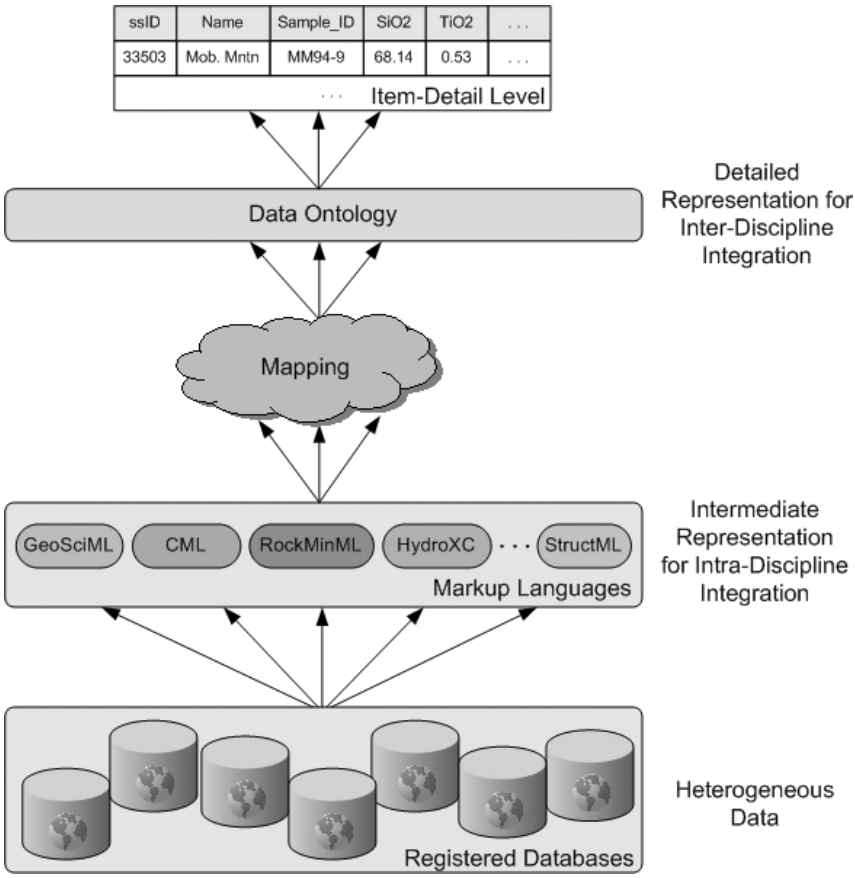

Fig. 2. Using Markup Languages and Ontologies for Integration — Heterogeneous data items can be represented through Markup languages (if any), and/or mapped to defined ontology(ies) for detailed representation.

a Compound Material shown in Fig. 3 would be ontologically linked to the concept of *Chemical Compound* in Fig. 4. Similarly, *lithology* (Fig. 3) would be mapped to the concept of *Rock* (Fig. 4), *mineral name* (Fig. 3) would be mapped to a *natural mineral* (Fig. 4), etc.

2.2. *Inter-discipline integration through ontologies*

As the Web evolves from a set of isolated application systems to a network of interacting disparate systems, the need to represent the semantics of the exchanged information such that it could be automatically understood is becoming a necessity. This is where *ontologies* can play an integral role. Ontologies aid in providing machine processable semantics of the information communicated between heterogeneous systems [45]. In ontologies, the semantic description of data, i.e., the logical relationships between data elements and other formal statements are made explicit

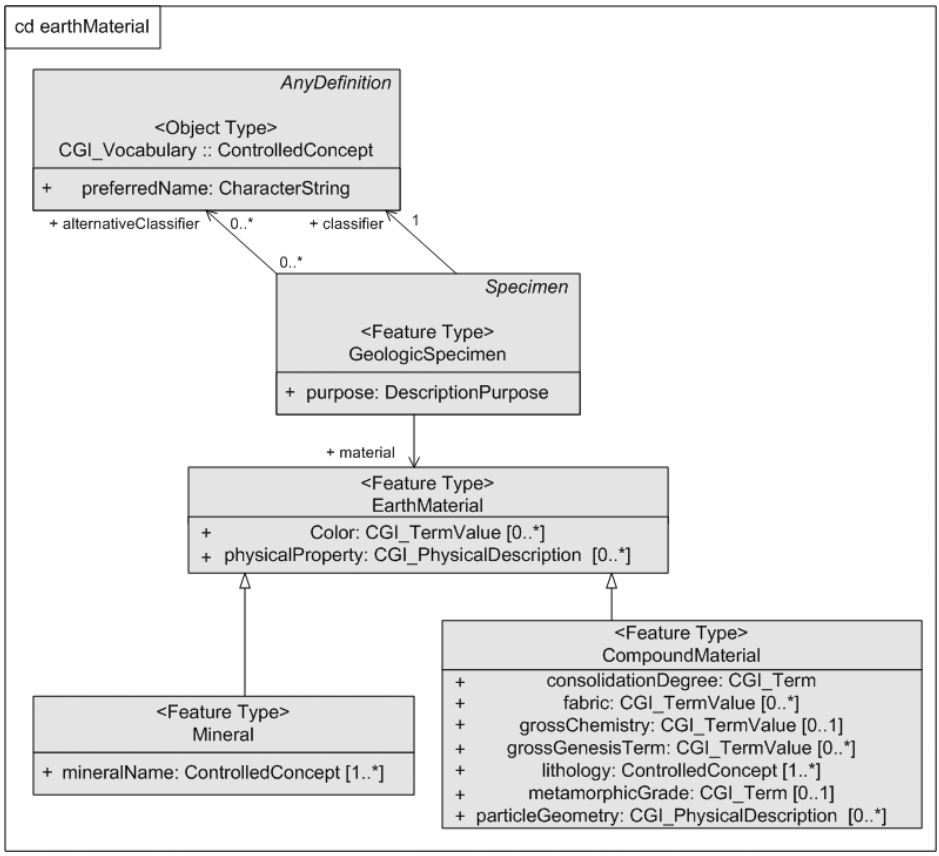

Fig. 3. Earth material package as defined in GeoSciML.

through ontology languages such as OWL (Ontology Web Language). This not only makes it easier for other human users of models to understand the specifically intended meaning, but also means that other tools can use the definitions transparently [17]. Thus, an ontology can be defined as a knowledge representation, different in part from an ML schema which mainly defines a message format [20].

To enable the sharing and integration of data on a global scale, along with other researchers we have introduced the idea of ontology-based data registration and discovery in geosciences [51, 75]. Our goal in defining ontologies is to provide an organizational structure for classifying data that can be discovered automatically by computers [48]. These high-level ontologies and object level ontologies, allow geologists to discover databases as well as datasets within databases using geoscience related concepts instead of simple keyword-based search. This is made possible as ontologies allow the registration of databases at different levels of granularity. For example, a relationship between occurrence of ignimbrites and hazardous volcanic eruptions can be inferred by an automated reasoning system even though this fact is

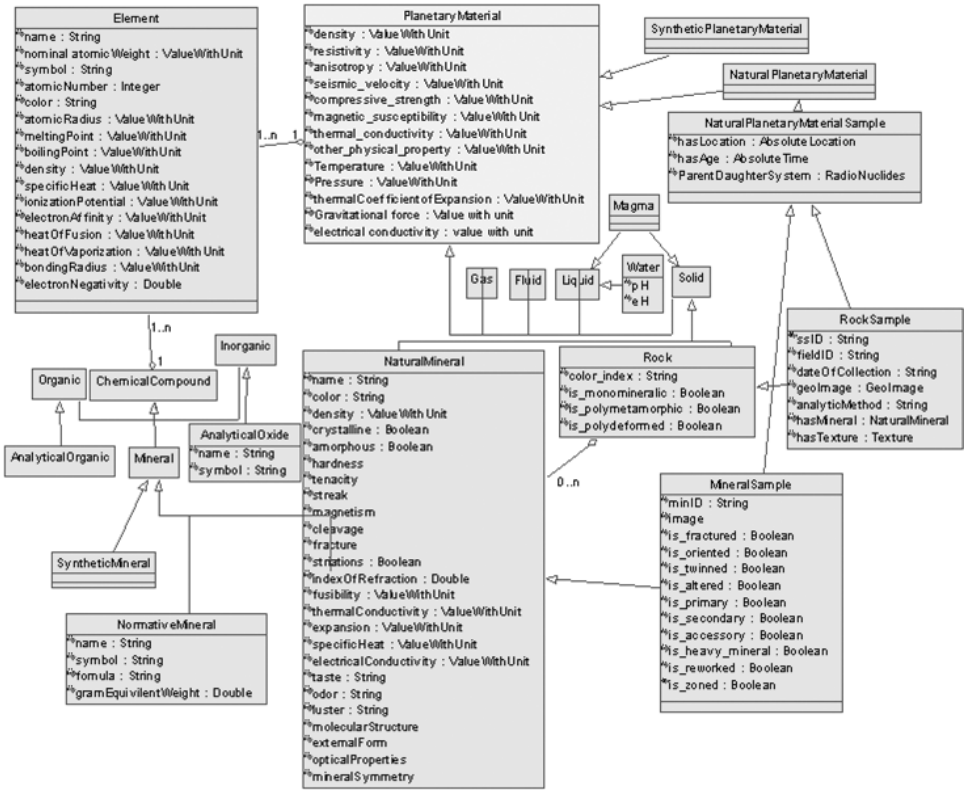

Fig. 4. Snapshot of EPONT: *Planetary Material* shows concepts at item detail level.

not contained in the database, if the ontologic framework effectively captures such a relationship. As shown in Fig. 5, a logical conclusion can be achieved about the relationship between ignimbrite and hazardous volcanic eruptions. The conceptual relationships are based on:

(1) Ignimbrite *is a* pyroclastic rock *is a* volcanic rock *is a* rock.
(2) Hazardous eruption *is an* Explosive eruption *is an* eruption.
(3) Explosive eruption *has Material* pyroclastic rocks.

Therefore, ignimbrites are a product of hazardous volcanic eruptions.

As mentioned earlier, the use of these ontologies, especially at the most detailed level, facilitates semantic integration of heterogeneous geologic datasets. Integration of databases can be done through (1) schema merging, when the user is knowledgeable about the organization (semantics of the schema), (2) view based techniques which include the creation of a virtual schema to allow the user to address structural heterogeneity, and (3) ontology based integration accomplished by registering databases to ontology. We favor ontology based integration as it

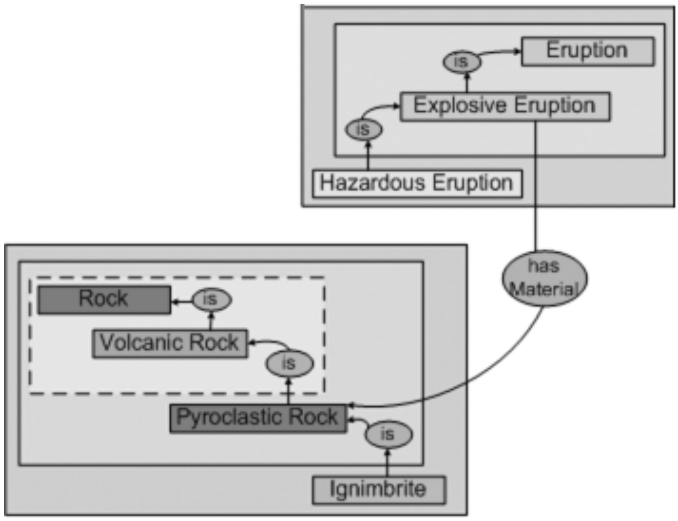

Fig. 5. A graphical representation of an ontology leading to automated capability of making a logical deduction through defined taxonomies and inference rules.

systematically resolves both syntactic and semantic heterogeneity, thus allowing integration of multiple distributed databases. Unlike integration based on merging multiple schemas, ontologic registration relates the data to concepts, rather than the structure within a database [41]. This facilitates queries over different levels of ontology classes and their instances, e.g., more specific/general classes can be found, or individual tuples matching a given query can be found [28].

3. Semantic Registration and Discovery

We have built a semantics-enabled service-oriented computational infrastructure that enables scientists to Discover, Integrate, and Analyze earth science data through all levels (upper, mid, and domain) and classes (object, process, and service) of ontologies. The resulting "engine" is called DIA (Discovery, Integration, and Analysis engine). The architecture of DIA has been designed to access and utilize classes of ontologies mentioned earlier to enable discovery, and numerical solutions to queries. The current implementation of DIA supports mid and domain level object ontologies, with the other ontology classes under development. It is designed to provide earth scientists the capability to better understand the relationships between the observed geologic records and the complex processes that have shaped them over the years. As its name suggests, DIA comprises three main phases: discovery, integration, and analysis.

Data discovery enables the users to retrieve the distributed data sets, i.e., located at multiple sites that are pertinent to the research task at hand. Although the engine currently uses object ontologies for discovery of data, it can be extended to access

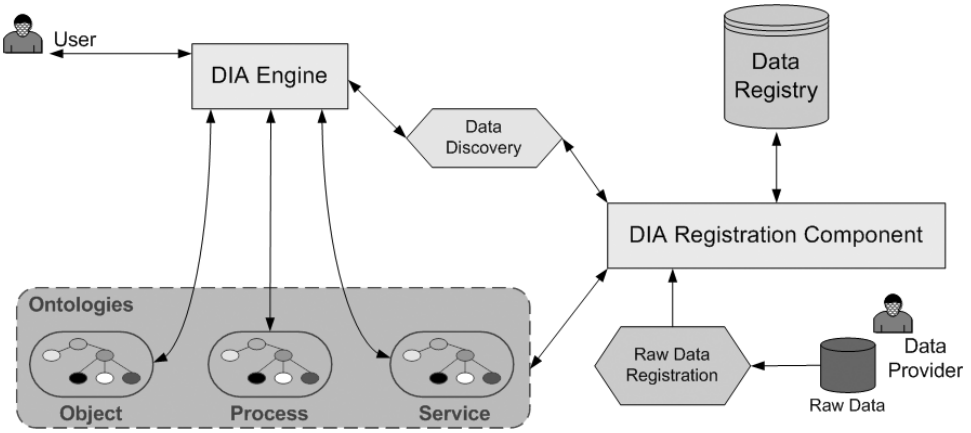

Fig. 6. Conceptual organization of classes of ontologies within DIA. The registration component is shown as a stand-alone component for clarity only.

data registered through markup languages. As mentioned earlier, object ontologies and markup languages define essentially the same concepts through similar classification structures, with the difference being the extra reasoning capabilities available in ontologies.

In DIA, users query various data sets along some common attributes to extract previously *unknown* information called "data products". The data products that are generated can either be used in their delivered form or used as input to the data analysis phase. Data analysis may be used to verify certain hypothesis or it may refine the data product with further data discovery and integration. Such a cyber-infrastructure for the geosciences is a requirement for both improved efficiency and trust for online conduct of science.

To illustrate the different DIA phases, consider the following example query: "Find the chemical composition of a liquid derived by 30% partial melting (PM) based on the average abundances of Rare Earth Elements (REE) of A-Type plutons in Virginia, and identify the correlation between these plutons and their geophysical (e.g., gravity) properties."

A number of steps are involved in answering the above mentioned query. These are: discovering data, identifying the A-Type bodies in VA, computing the averages, using the REE definitions contained in the element ontology and exporting the data to a PM tool for computation, deploying gravity modeling tools after accessing gravity data [75], and displaying the results. Note that the above stated query uses both object and process ontologies (e.g., A-Type bodies data resides with an object ontology while the partial melting concept and associated numerical models are contained in a process ontology). We will trace these steps through the DIA engine, which entails five phases: (i) resource registration, (ii) query specification, (iii) resource discovery, (iv) filtering and integration, and (v) analysis.

3.1. *Resource registration*

DIA's registration system (denoted SEDRE) allows geoscientists to register their data (which is normally in *Microsoft Excel* files) to one or more ontologies, for the purpose of sharing. As mentioned earlier, since we have not used MLs in constructing DIA, only data-ontology mappings are carried out in the registration phase. The goal is to allow researchers to associate one or more ontologies to their files so that a unique and definite meaning is associated with each column. Ontological registration allows relating a column to columns with similar (or close) semantics in other files. As markup languages also provide similar capabilities, incorporating MLs in DIA would allow access to data through ML format, and "advanced" querying over the data through ontologies. The semantic annotation of ML's, and ML to ontology mapping are active research topics and are beyond the scope of this paper. The interested reader is referred to [8, 63]. SEDRE facilitates discovery through resource registration at three levels:

(i) *Keywords-based registration*: Discovery of data resources (e.g., gravity, geologic maps, etc.) requires registration through the use of high level index terms. For instance, the popular AGI Index terms (or an extension thereof) can be used. These terms and their extensions can then be cross indexed to other indexes such as GCMD and AGU.

(ii) *Ontological class-based registration*: Discovering item level databases requires registration at data level ontologies. In this case, a data set may be registered to a data level ontology, e.g. bulk rock geochemistry, gravity database, etc.

(iii) *Item detail level registration*: Item detail level or fine-grain registration consists of associating a column in a database to specific concept or attribute of an ontology, thus allowing the resource to be queried using concepts instead of actual values. This mode of registration is most suitable for datasets built on top of relational databases. However, item detail level registration can be extended to cover Excel spreadsheets and maps in ESRI Shapefile format by internally mapping such datasets to PostgreSQL tables. For example, a column in a geochemical database may be specified as representing SiO2 measurement. This level of registration is a requirement for semantic integration, i.e., the automatic processing (by tools) of shared data.

Figure 7 shows the schematic representation of data registration through SEDRE and discovery/integration through DIA. The three heterogeneous data sets discussed above are registered using the defined semantic data ontologies. The different terms in the individual data sets are mapped to the terms defined in the ontology. For instance, SO_2 columns in the data sets are mapped to respective terms in the ontology. SEDRE allows the data owners to maintain control over their data (if they wish to), and in this case only store the data — ontology terms mappings. The mappings are tagged using longitude/latitude coordinates to enable efficient access to relevant data. We recognize that data registration through ontologies is a time

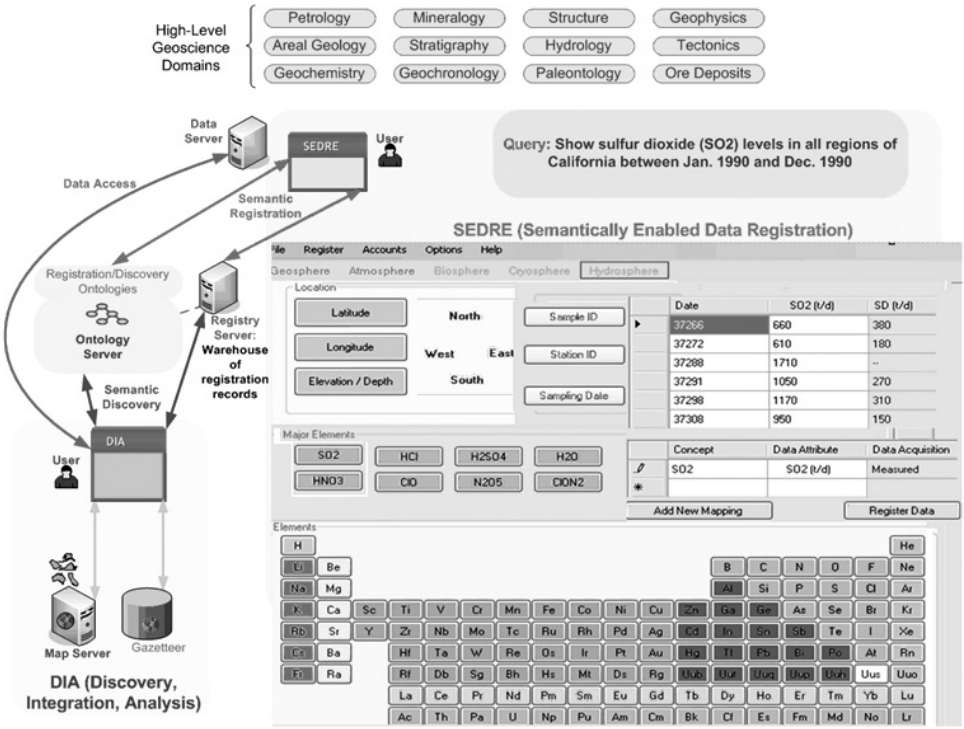

Fig. 7. Schematic representation of registration of data through SEDRE and discovery/integration through DIA. Using the defined ontologies, users can register/discover the data. SEDRE's interface allows easy tagging of data items to ontology terms. For instance, SO_2 columns in the data sets are mapped to respective terms in the ontology. The mappings are also tagged using longitude, latitude, elevation, etc. coordinates to enable efficient access.

consuming process, and that data owners may not be able/willing to register their data sets in "one go". Therefore, SEDRE is developed as a downloadable service, where data owners can download SEDRE (along with all the required ontology terms) on their private machines, and connect to SEDRE's online repository only to upload the data-ontology mappings. This allows data owners to register their data at their own convenience, while keeping ownership of data. DIA uses different "Registry Servers" (RSs) which could be distributed worldwide, to provide directory functionalities (registration of data and tools, indexing, search, etc.) The providers of resources advertise their resources on registry servers, which may then be (manually or automatically) discovered and used.

3.2. *Query specification*

Currently, the DIA engine supports the user query to be expressed in a menu-based format. This lets the user to select only specific items, which in turn queries only a subset of the data. The user does not need extensive knowledge of the querying techniques, models or keywords (which may be required in a text-based format).

The task at hand can be completed with the help of a few "mouse-clicks," as the user clicks through the different menus to "build" an exact query.

The "User Interface" (UI) component of DIA is an *ArcGIS Server / .NET* map viewer Web application. Thus, currently only two-dimensional geometries are supported. However, UI is an extendable component and higher-dimension geometries could also be supported. DIA uses two Web servers. The first Web server is responsible for routing users' queries to DIA's query processor and the second ensures communication between DIA's query processor and its own map server. In a minimal deployment, a single Web server may be used for both purposes (we use a single instance of Windows IIS Web server as DIA's Web server). Map-based queries can be refined by specifying a bounding box that identifies a pair of latitude-longitude points which delimits the query's spatial scope. After the query's spatial scope is specified, the user uses DIA's drop-down menu to indicate the filters (A-Type igneous rock filter in our running example) and/or tools to be applied to the data samples discovered in the query's spatial scope. Figure 8 shows what menus are involved in getting to the A-Type filtering (four top-level menus are navigated to get to the A-Type menu). Similarly, the "Tool Selection" layer in Fig. 9 shows how menus are navigated to invoke the required tools.

In DIA, the "Query Processor" (QP) module is responsible for producing the results for users' queries and delivering them to the Web server. The QP consists of two sub-components: (i) the query interpreter and (ii) the geology and mapping filters and tools. The former is a *.NET* module that interprets queries and identifies the appropriate filters and/or tools to be invoked to answer each query. The latter is a large set of *.NET* modules that perform DIA's core functionalities including filters (e.g., A-Type igneous rock filter), tools (e.g., kriging) and map management routines (e.g., coloring of geological bodies and sample points.

3.3. *Resource discovery*

In resource discovery, the DIA engine identifies and retrieves the resources (data and tools) required to answer the user's query. When the QP receives the query from the Web server, it determines the type of data required to answer the query. In our running example of A-Type plutons, the QP determines that data associated with the keyword "GeoChemistry" is the query's target. The QP then interacts with one or several registry servers to retrieve the needed data. An example of a registry server is available at the GEON server (www.geongrid.org). To interact with

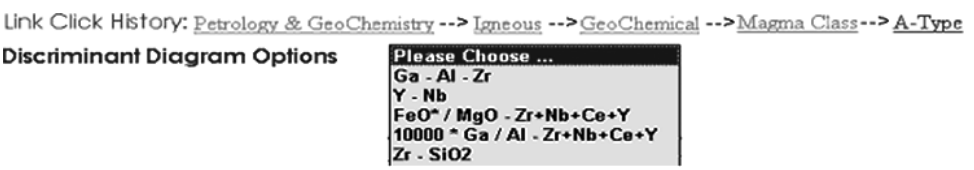

Fig. 8. Query specification through menus. Menus are dynamically generated as per the defined ontologies, and available tools/services.

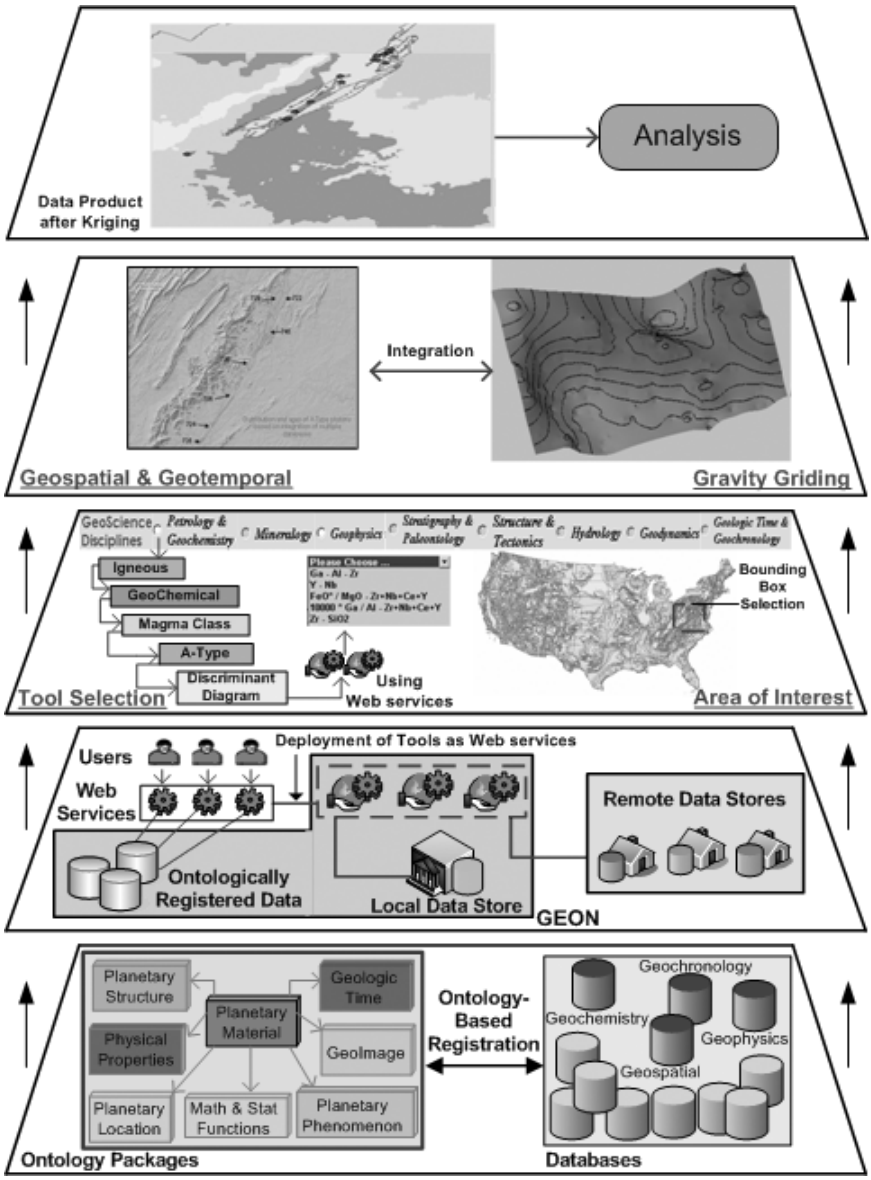

Fig. 9. A layered bottom-up depiction of DIA phases from data to data products — The figure shows the functioning of DIA in five layers, starting from the basic constructs of the system that include the ontologies, markup languages, and the raw data from various disciplines in different databases. The data is registered according to the defined ontologies and markup languages. The second layer shows that the registered data (local or remote) can be semantically discovered and used by Web services that represent various tools required by geoscientists. The third layer shows that these tools (wrapped as Web services) can be discovered by navigating through different menus, which are generated according to the defined ontologies and available tools. The layer also shows that using a bounding box, a user may define his/her "area of interest". Based on the user query (identifying A-Type plutons and their ages in our case), DIA outputs the result on the map, as shown in the fourth layer. This layer also shows that gravity data for the area of interest is then integrated with the previous result, leading to the data product shown in layer five (on which user analysis can be performed).

the GEON server, DIA invokes a GEON Web service called GEONResources that provides functions for searching and getting the metadata information for resources registered through GEON portal. When invoking GEONResources, DIA's QP indicates that it is searching data sets registered with the keyword "GeoChemistry" and that contains data samples in the query's spatial bounding box. For each returned database, the DIA system executes a two-step process. First, it builds a virtual query expressed in *SOQL* (a language developed by GEON researchers at SDSC) that requests all the data (i.e., columns) that are necessary to apply the filter specified by the user. The DIA system then invokes a GEON Web service called SoqlToSql that translates this *SOQL* query into an *SQL* query. In the second step, DIA submits the *SQL* query to the GEON server that interacts with the actual database server, gets a record set containing the relevant data samples, and returns the data to the DIA engine.

3.4. *Filtering and integration*

Data filtering is a process in which DIA engine transforms raw data into a data product. After DIA retrieves the data sets relevant to the user's query, it determines whether the filter(s) to apply or tool(s) to use is locally available. If so, the filter/tool is applied to the data sets and the query result is displayed to the user. If not, DIA searches for the needed filter/tool in registry servers. DIA is able to invoke any external tool that is wrapped as a Web service (similar to invoking the GEON Web service above, for resource discovery). In the case of the given A-Type query, the A-Type filter is already available in DIA and also made available as a Web service for external users.

Integration in DIA is a process in which the results of several sub-queries are produced and then presented through the map-based user interface. There are two main classes of integration: *Intra-class integration* is a process whose input is two or more homogeneous data sets, i.e., registered to the same ontology (or defined using a markup language if one is used). This process uses the common ontology to interpret all data sets and generate an integrated data product. *Inter-class integration* is a process whose input is two or more heterogeneous data sets (i.e., registered to different ontologies.) This process uses the appropriate ontology to interpret each data set. It uses an integration class to generate a data product out of two or more data sets.

In our running example, to produce the result of gravity kriging data, DIA follows the same workflow as shown above for determining A-Type bodies. DIA looks up registry servers for gravity data in the selected area of interest and then retrieves the raw gravity data from its provider(s) (e.g., http://paces.geo.utep.edu). DIA then determines whether a gravity kriging tool is locally available. Since such a tool is already included in DIA's implementation, it is invoked and no external registry servers are searched. When the output of the kriging tool is generated, DIA overlays it on the previously generated results (i.e., A-Type plutons) making it possible for

the user to have a natural and easily interpretable view of the integration's result. Figure 9 shows the integration between geospatial, geotemporal, and gravity griding data products that are obtained after data filtering through a "bounding box" (used to limit the area of interest).

3.5. *Analysis*

Geoscientists can use the data products generated as a result of the above mentioned phases in hypothesis evaluation, i.e., to analyze the results. In Fig. 9, using a space-time analysis and studying the generated data product after the kriging, a geoscientist can verify the hypothesis on A-Type bodies. Our running example query (*A-Type rocks*) has major implications on tectonic models such as (1) a failed rift associated with a triple junction [60], (2) gravitational collapse of crystal regions over thickened by Grenville orogenesis [71], (3) flanks of an active within plate rift zone similar to the Red Sea region [76], and (4) as a continental plume track [26]. Thus, facilitating rapid data /tool access clearly is a requirement as geoscientists engage in more complex queries.

3.6. *DIA's service-oriented approach*

The DIA engine (Fig. 10) is a Web-based, service-oriented system developed using a variety of technologies including: ESRI's ArcGIS Server 9.1, Microsoft's .NET framework, Web services, Java, and JNBridge 3 (facilitates communication between .NET and Java). Users submit queries through the DIA's Web-accessible graphical interface. The engine translates these queries into a sequence of tasks that include: accessing map servers, discovering and accessing data sources, invoking Web services, filtering features, joining layers, and the graphic rendering of query results for visualization. Currently, data discovery and access is limited to object ontologies in DIA. However, DIA's extendable design allows process and service ontologies (future development) to be incorporated. Similarly, markup languages can also be used in tandem with object ontologies to facilitate data discovery and exchange. Figure 10 shows how both markup languages and ontologies can be used in tandem to register and serve data. The DIA engine also enables users to save their query history as well as export data products for future references. Since the DIA engine is developed along a service-oriented approach, key code modules are wrapped as Web services. This approach has two advantages. First, it makes the system readily extensible. As the geoscience community introduces new services, these could be integrated in the DIA engine as new functionalities. Second, services developed for the DIA may be used as building blocks to produce other systems.

To illustrate DIA's functionality, consider the following query that was introduced in Sec. 1:

Q1: "*What is the distribution of U/Pb zircon ages of A-Type plutons in Virginia. Identify the correlation between these plutons and their geophysical* (*e.g., gravity*) *properties*".

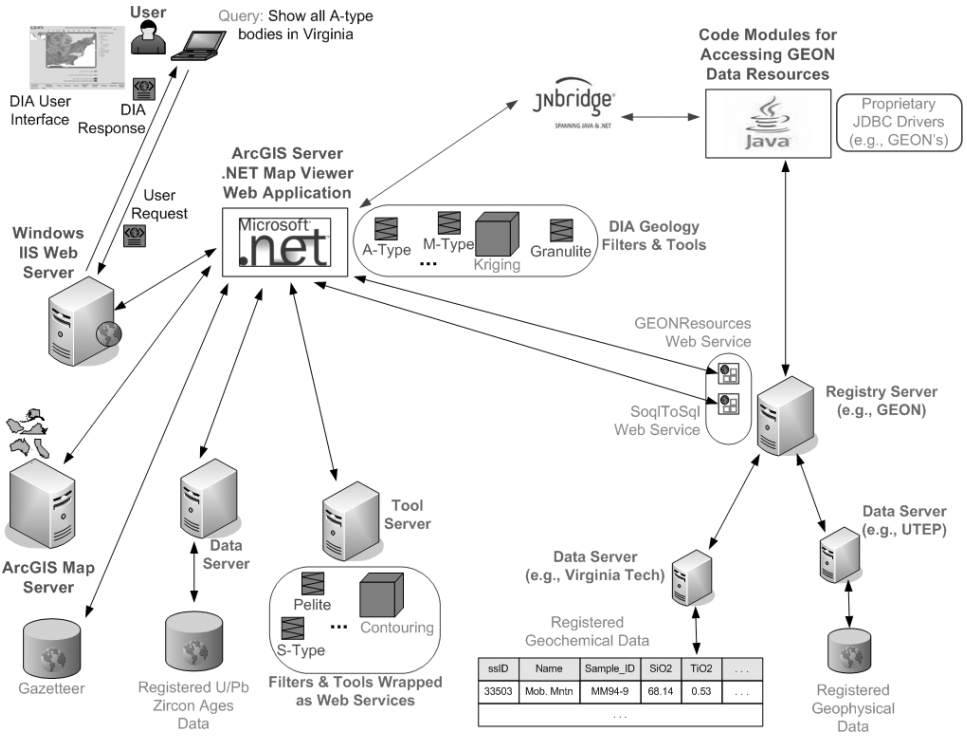

Fig. 10. DIA engine architecture.

In the following, we annotate the different "Steps" that DIA goes through in answering the query, for better explanation.

Step 1: DIA's browser-based interface presents two querying options: geologic-map-based and region-based querying. Region-based queries provide answers to high level queries about specific regions. On the other hand, geologic-map based queries focus on more "thorough" rock records and are more data-intensive. In this example, we use the geologic-map based querying.

Answering Q1 requires four steps: selecting the A-Type filter to differentiate between plutons (Step 2), selecting the spatial scope (only look in VA) where the A-Type filter is to be applied (Step 3), invoking the kriging/contouring tool to identify geophysical properties (Step 4), and selecting the spatial scope (VA) where the kriging/contouring tool is to be applied (Step 5).

Step 2: As mentioned earlier, the different geoscience tools are available in the form of drop-down menus. These menus are located in the bottom of the screen, just below the geologic map. The user has to navigate the menus until he locates the desired tool to apply (in this case, the A-Type filter tool). The menus are based on the terms contained in EPONT. For example, A-Type rocks are a "Magma Class"

which is a "Geochemical Analysis" of the "Igneous rock" type, and finding plutons of a certain type is associated with the rock's "Petrology and Geochemistry". Therefore, the user navigates the menus by first, clicking on the radio button for "Petrology & Geochemistry". DIA then displays a menu containing three options for the three types of rocks: Igneous, Metamorphic, and Sedimentary. In this query, the user selects the "Igneous" option. In the next sub-menu, a number of options for performing different analyses are provided. In this case, the user selects the option that corresponds to "Geochemical analysis". Similarly, in the next menu, "Magma Class" option is selected. This displays a four-option menu for the A, I, S, and M-type of rocks. The user selects the option "A-Type". Since DIA supports several A-Type filters, it will display a drop-down menu of the available filter tools. For example, if you select the option Ga-Al-Zr, DIA will use the A-Type filter tool based on the approach using elements Ga, Al, and Zr proposed in Whalen *et al.*, 1987.

Step 3: Now that the A-Type filter is specified, the user needs to select the specific area of interest, i.e., the state of Virginia in Q1. To accurately specify Q1's spatial scope, the user can zoom in the geologic map. Then, the user can draw a bounding box around VA where the A-Type plutons are to be found. DIA will then compute all the A-Type plutons in the selected region and display them along with the samples used in the computation as shown in Fig. 11.

Step 4: To generate the gravity map of the selected area, the user needs to return back to DIA's main menu. This is facilitated by maintaining a "Link Click History". At the top menu, instead of "Petrology and Geochemistry", the user can now select "Geophysics" (another Geoscience discipline). DIA will then display sub-options. The user can click on the option "Gravity". This will display a drop-down sub-menu related to tools available for Gravity analyses. To generate the gravity map, the user can select the option of "Kriging/Contouring".

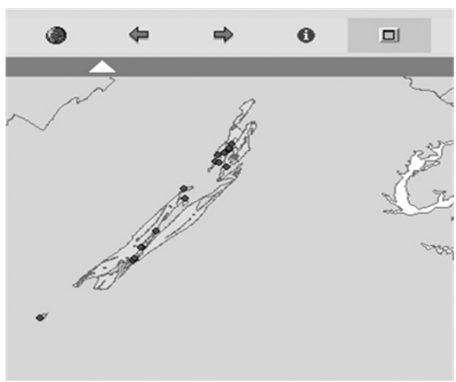

Fig. 11. A-Type Plutons in a region in Virginia.

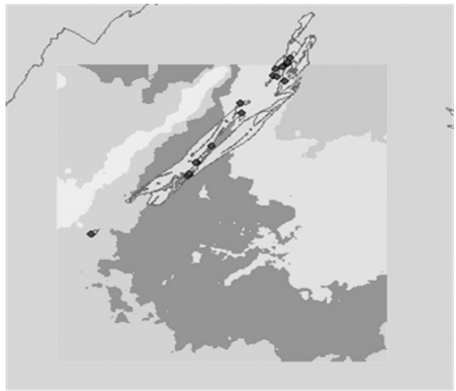

Fig. 12. DIA's answer to Q1: A-Type plutons and their geophysical properties.

Step 5: In the last step, the user needs to specify the area of interest as previously done for the A-Type plutons, i.e., using the bounding box button to indicate the scope for the Kriging tool. The final data product that DIA generates in response to Q1 is shown in Fig. 12. The user can click on the discovered A-Type plutons to get their U/Pb ziron ages.

3.7. *Factors affecting DIA's performance*

Since DIA is designed as a distributed system, there are a number of potential bottlenecks that may affect its performance. In this section, we provide a brief overview of the performance issues for DIA in its present form.

Starting with the semantic registration of data, DIA allows data owners two options: (i) register their data along with the ontologic mappings, or (ii) register only the data-to-ontology mappings. Since the Registry Server will remain unresponsive for the duration of the upload, the amount of registration data will directly influence DIA's performance (currently a single Registry Server is employed). Therefore, in the first case, small to medium datasets provide minimal impact, while large datasets (in orders of Giga, Tera, etc. bytes) inhibit the system's performance greatly. These performance issues may be handled in an ideal system through replication, separation of server concerns (register vs. serve), etc. In the second case, DIA's performance is tied directly to the data owner's (e.g., GEON and UTEP in previous sections) ability in responding to DIA's data requirement.

Users interact with DIA through a map, and nested (hierarchical) menus that are populated as per the defined ontology. Since the Map server has to communicate with the query processor through another Web server, the delivery of maps is the major bottleneck affecting system performance. Menu generation is not greatly time consuming for conditions where only the local ontology is concerned, i.e., traversing the ontology and providing navigation options through menus. In cases

where menus have to be populated according to the available tools, performance may vary. For instance, in our running example, A-Type classification tools are required. DIA searches for both "in-house" and external options made accessible through Web services. Currently, Web services-based resources are tightly coupled so performance is predictable. However, if a Web services registry (like UDDI) is consulted, it will likely affect DIA's performance. In our running example of A-Type plutons, the GEON server is consulted for data. This is done by first invoking a Web service called GEONResources to search and retrieve the metadata information for resources registered through GEON portal. Then, a *SOQL* query is built, and another Web service called SoqlToSql that translates this *SOQL* query into a *SQL* query is invoked. In the final step, DIA submits the *SQL* query to the GEON server that interacts with the actual database server, and returns the data. DIA's performance thus depends directly on GEON, which is itself dependent on the actual database server (which may lie outside GEON). In addition, the SoqlToSql Web service is known to DIA, but if a new Web service is to be located to perform the required transformation, that adds to DIA's response time.

Since DIA's output is map-based, its performance for integration results can be gauged by verifying the retrieved samples. In our running example, to produce the result of gravity kriging data, DIA exhibits same performance issues as mentioned above for determining A-Type bodies. Gravity data is first searched, a gravity kriging tool is then identified, and the output of the kriging tool is generated by overlaying it on the previously generated results (i.e., A-Type plutons) on the map. Like any distributed system, the complexity of the query will thus directly influences DIA's performance.

4. Related Work

In this section, we provide a brief overview of major research efforts that are closely related to the approach proposed in this paper. [9] investigates the development of categories shared in the field logging of a region by a team of geologists. It uses visualization, neural networks and spatial statistical tools employed to analyze the complex space of attributes observed and the categories developed. The study suggests the use of contextual factors to deal with category discrepancy that exist between individuals and when adopting ontological approaches to information representation. The Global Change Master Directory's (GCMD) earth science ontology is limited to keyword lists for classification [53]. NASA's Semantic Web for Earth and Environmental Terminology (SWEET) ontologies are essentially class hierarchies with some limited expression of properties. We have used these class hierarchies in defining EPONT. The FGDC (Federal Geographic Data Committee) Content Standard for Digital Geospatial Metadata [24] was developed in 1994 to describe all possible geospatial data. However, since the standard is very complex (with 334 different elements, 119 of which exist only to contain other elements), its adoption and use has been limited.

In earthquake science, [15] proposes a semantics-based system to improve interoperability among heterogeneous earthquake data. The Earth System Grid (ESG) project [58] aims to provide discovery of large datasets based on grid technologies and the use of metadata schemas and prototype ontology. [47] introduces two use cases within the e-Wok Hub project: documentary search and subsurface modeling. It describes a knowledge-driven methodology based on semantic annotation that can be used in both cases. It shows that the definition of domain ontologies and the development of a semantic annotation methodology enable to identify and handle raw data, geological objects and geological interpretations according to their semantic contents. [65] describes an approach to ontology design called Workflow-Driven Ontologies (WDO). WDOs capture basic knowledge about a domain. Use cases typically drive the specification of domain-based ontologies. A case study from earth science is used to illustrate WDO. Abstract workflow specifications drive the elicitation and specification of classes and their relationships. For example, earth scientists start the knowledge acquisition process by identifying a product and identifying methods that can generate the product. Then, they identify data that are required as input for the identified methods. Earth scientists can refine WDOs by refining a WDO-derived workflow. Related efforts are underway to enable semantic and interoperable geospatial and geographic systems [21, 27].

The Geospatial Semantic Web Interoperability Experiment (GSWIE) [30] brought together a number of threads in semantics, Web technology, and geospatial processing. The principal areas of this experiment include the development of geospatial, domain, and other ontologies covering the knowledge and operations domains of the demonstration use case; the development of a reference Web architecture for exchanging information with formal semantics and processing information queries; and the choice and refinement of one or more query languages and predicates for expressing geospatial queries. [38] introduces five ontology types in OWL that contribute to forming a geospatial semantic system and discuss their use within GSWIE. The base geospatial ontology provides the core geospatial knowledge vocabulary and knowledge structure; the feature data source ontology provides an ontological view of WFS (Web Feature Service Implementation Specification) data; the geospatial service ontology enables knowledgebase discovery and execution of all registered geospatial services; the geospatial filter ontology enables the integration of geospatial relationships into the queries; and the domain ontology provides a knowledge representation that is organized, customized, and aligned with a specific domain and/or user.

SWING (Semantic Web Services INteroperability for Geospatial Decision Making) investigates the applicability of semantic technologies in the area of geospatial services [4]. It leverages semantic Web service standards such as Web Service Modeling Ontology (WSMO) [78] and the Web Service Modeling Language (WSML) [23] for the semantic annotation of geospatial services to increase the efficiency and accuracy of discovering and integrating geospatial services. [42, 43] defines an

ontology-based approach for the discovery of geographic information (GI) services. It uses ontologies for describing geospatial operations and creating descriptions of requirements and service capabilities. Then, it describes matches between these descriptions based on function sub-typing. [78] proposes two semantic Web enabled geospatial frameworks. The first framework integrates rules and ontologies for expressing and reasoning over symbolic geographic knowledge. The second framework is a hybrid extension of the basic framework with geospatial information processors that are more suited to manipulating the geometrical (location) component of the information. [25] proposes a framework that allows the mapping of a spatial ontology and a geographic conceptual schema. The mapping of ontologies to conceptual schemas is made using three different levels of abstraction: formal, domain, and application levels. At the formal level, highly abstract concepts are used to express the schema and the ontologies. At the domain level, the schema is regarded as an instance of a generic data model. The application level focuses on the particular case of geographic applications. [56] proposes a semantic metadata management system based on ontologies and use of Semantic Web languages such as OWL [35]. The system defines an ontological data model for providing the spatial, temporal, presentation, distribution, and identification properties of data. Moreover, a data content class is defined that uses actual domain concepts defined in the geoscience ontology. However unlike our approach, no such ontology is discussed/defined.

5. Conclusion

We have presented an approach for the semantic integration of data and tools for geosciences. We propose that markup languages be linked to mid-level ontologies for a more comprehensive understanding of the meaning of data leading to integration, and other classes of ontologies (process and service) be developed to facilitate knowledge creation. As an initial step towards taking geosciences to the envisioned Web (a.k.a., semantic Web), we have developed the DIA engine. DIA uses ontologies and Web services to organize, annotate, and define datasets and geoscience tools. DIA is designed mainly as a proof-of-concept, and its extensible design allows the incorporation of our defined vision, i.e., using various classes of ontologies with markup languages developed for different geoscience disciplines.

We expect that as the semantic Web matures, and more geoscientists adopt the service-oriented paradigm, a number of geoscience tools will be made accessible as Web services. This would require that similar to data management through ontologies and markup languages, Web services for tools be also registered to a "service ontology". Annotating Web services with semantics would ensure that appropriate tools (in form of Web services) are selected in an efficient and automatic manner for answering geoscience queries. Domain experts would provide formal specifications of geoscience concepts, enabling automated Web service usage. Moreover, since the semantic Web is geared towards interactions involving minimal human-intervention,

service ontology would enable direct service-to-service communication, automated reasoning, and ease of information transfer. Thus, service ontology will do for Web services what data ontology has done for geoscience data. We believe this will prove to be a major step in taking geoinformatics further.

Appendix A. Glossary of Selected Terms

AGI: American Geological Institute
AGU: American Geophysical Union
CGI: Commission for the management and application of Geoscience Information
DIA: Discovery, Integration, and Analysis Engine
DOLCE: Descriptive Ontology for Linguistic and Cognitive Engineering
EPONT: Earth and Planetary Ontology
ESRI: Economic and Social Research Institute is a software development and services company providing Geographic Information System (GIS) software and geodatabase management applications
GCMD: Global Change Master Data Directory
GIS: Geographic Information System
GEON: Geosciences Network project aims to develop cyberinfrastructure in support of an environment for integrative geoscience research.
GeoSciML: GeoScience Mark-up Language
Integration through layering: Overlay of data products as commonly utilized in GIS methods
Integration through semantics: Semantic Integration is a set of technologies drawn from Artificial Intelligence, Linguistics and Knowledge Management designed to help make sense of complex information and allow improved integration between systems
ML: Markup language = A notation for identifying the components of a document to enable each component to be appropriately formatted, displayed, or used. A markup language, e.g., XML provides a way to combine a text and extra information about it
NADM: North American Geological Data Model
Ontology = A set of knowledge terms, including the vocabulary, the semantic interconnections, and explicit rules of inference and logic for some particular topic (www.ontology.org), (Gruber, 1993)
OWL = Web Ontology Language is a family of knowledge representation languages for authoring ontologies
PM: Partial Melting is a geological phenomenon
REE: Rare Earth Elements
Registration = Process of adding new descriptions to a repository
SEDRE: Semantically Enabled Data Registration Engine
Service registry = Is a network-accessible directory that contains information about the available services

SUMO: Suggested Upper Merged Ontology
SWEET: Semantic Web for Earth and Environmental Terminology
UML = Unified Modeling Language
UTEP: University of Texas at El Paso
XML: Extensible Markup Language
XTM = This specification provides a model and grammar for representing the structure of information resources used to define topics, and the associations (relationships) between topics.

Acknowledgments

This research project was funded in part by the National Science Foundation Grant EAR-022558 to A. Krishna Sinha, and IIS-0916614 and IIS-0811954 to Mourad Quzzani. We also appreciate thoughtful discussions about ontologies with Kai Lin, Cal Barnes and Boyan Brodaric.

References

[1] A. Abdelmoty, P. Smart, B. El-Geresy and C. Jones, Supporting frameworks for the geospatial semantic web, *Proceedings of the Symposium on Spatial & Temporal Databases*, LNCS 5644, Springer, 2009.
[2] K. Aberer, Special issue on peer to peer data management, *SIGMOD Record* **32**(3) (2003).
[3] B. Amann, C. Beeri, I. Fundulaki and M. Scholl, Ontology-based integration of XML web resources, *International Semantic Web Conference '02*.
[4] M. Andrei, A. Berre, L. Costa *et al.*, SWING: An Integrated Environment for Geospatial Semantic Web Services. The Semantic Web: Research and Applications, *5th European Semantic Web Conference, ESWC 2008*, 2008, pp. 767–771.
[5] F. Baader, I. Horrocks and U. Sattler, Description logics as ontology languages for the semantic web, In D. Hutter and W. Stephan (eds.), *Mechanizing Mathematical Reasoning: Essays in Honor of Jörg Siekmann on the Occasion of His 60th Birthday, Number 2605 in Lecture Notes in Artificial Intelligence, 228–248* (Springer, 2005), pp. 228–248.
[6] F. Baader, D. Calvanese, D. McGuinness, D. Nardi and P. Patel-Schneider, *The Description Logic Handbook: Theory, Implementation, and Applications* (Cambridge University Press, 2003).
[7] T. Berners-Lee, J. Hendler and O. Lassila, The semantic web, *Scientific American* **284**(5) (2001) 34–43.
[8] H. Bohring and S. Auer, Mapping XML to OWL ontologies, *Marktplatz Internet: Von e-Learning bis e-Payment*. Leipziger Informatik-Tage (LIT2005), Leipzig, Germany, 2005, pp. 147–156.
[9] B. Brodaric and M. Gahegan, Learning geoscience categories in situ: Implications for geographic knowledge representation, *ACM-GIS* (2001) 130–135.
[10] A. Bouguettaya, Z. Malik, A. Rezgui and L. Korff, A scalable middleware for web databases, *Journal of Database Management* **17**(4) (2006) 20–47.
[11] T. Bray, J. Paoli, C. Sperberg-McQueen, E. Maler and F. Yergeau, *Extensible Markup Language (XML) 1.0 (Fourth Edition) — Origin and Goals* World Wide Web Consortium, Edited September 2006, Retrieved April 2007.

[12] C. Bussler, Semantic B2B integration, *ACM SIGMOD Record Journal* **30**(2) (2001) 625.
[13] D. Carlson, Semantic models for XML schema with UML tooling, *2nd International Workshop on Semantic Web Enabled Software Engineering* (*SWESE 2006*), 2006.
[14] F. Casati and M. C. Shan, Dynamic and adaptive composition of e-service, *Information Systems* **26**(3) (2001) 143–163.
[15] A. Y. Chen, A. Donnellan, D. McLeod *et al.*, Interoperability and semantics for heterogeneous earthquake science data, *Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*, Florida, USA, 2003.
[16] L. Chen, N. R. Shadbolt and C. A. Goble, A semantic Web-based approach to knowledge management for grid applications, *IEEE Trans. for Knowledge and Data Engineering* **19**(2) (2007) 283–296.
[17] Z. Cui, D. Jones and P. O'Brien, Semantic B2B integration: Issues in ontology-based approaches, *ACM SIGMOD Record Journal* **31**(1) (2002) 43–48.
[18] Y. Ding, D. Fensel, B. Omelayenko and M. Klein, The semantic web: Yet another hip, *Data and Knowledge Engineering* **41**(3) (2002) 205–227.
[19] H. H. Do and E. Rahm, COMA — A system for flexible combination of schema matching approaches, *Proc. 28th Intl. Conf. Very Large Data Bases*, 2002, pp. 610–621.
[20] J. Domingue and L. Cabral, OCML: Ontologies to XML schema lowering, AKT-SWS04, also available in the CEUR Workshop Proceedings, Vol. 122, December, 2004.
[21] M. J. Egenhofer, Toward the semantic geospatial web, *Proc. Intl. Conf. on Advances in GIS*, ACM, 2002.
[22] J. Euzenat and P. Valtchev, Similarity-based ontology alignment in OWL-Lite, *Proc. 16th European Conference on Artificial Intelligence*, 2004, pp. 333–337.
[23] D. Fensel and C. Bussler, The Web Service Modeling Framework WSMF, 2002.
[24] FGDC Metadata, http://www.fgdc.gov/metadata/metadata.html, August, 2010.
[25] F. T. Fonseca, C. A. Davis and G. Câmara, Bridging ontologies and conceptual schemas in geographic information integration, *GeoInformatica* **7**(4) (2003).
[26] M. Fokin and A. K. Sinha, Plume induced neoproterozoic magmatism in eastern laurentia and its bearing on the breakup of rodinia, *GSA Abstract with Programs* **43**, 2002.
[27] M. F. Goodchild, M. J. Egenhofer, R. Fegeas and C. A. Kottman (eds.), *Interoperating Geographic Information Systems* (Kluwer, New York, 1999).
[28] B. C. Grau, Y. Kazakov, I. Horrocks and U. Sattler, A logical framework for modular integration of ontologies, in *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI 2007)*, 2007, pp. 298–303.
[29] T. R. Gruber, A translation approach to portable ontologies, *Knowledge Acquisition* **5**(2) (1993) 199–220.
[30] GSWIE — Geospatial Semantic Web Interoperability Experiment, http://www.opengeospatial.org/projects/initiatives/gswie; Report OGC 06-002r1.
[31] A. Halevy, A. Rajaraman and J. Ordille, Data integration: The teenage years, *Proceedings of 32nd international Conference on Very Large Data Bases* (Seoul, Korea, September 12–15, 2006), U. Dayal *et al.* (eds), *Very Large Data Bases* VLDB Endowment, 9–16, 2006.
[32] J. Heflin and J. Hendler, A portrait of the semantic web in action, *IEEE Intelligent Systems* **16**(2) (2001) 54–59.
[33] J. Hendler and D. McGuinness, The DARPA agent markup language, *IEEE Intelligent Systems* **15**(6) (2000) 72–73.

[34] M. A. Hernandez, R. J. Miller and L. M. Haas, Clio: A Semi-automatic tool for schema mapping, *Proc. ACM SIGMOD*, 2001, p. 607.
[35] I. Horrocks, P. Patel-Schneider and F.-V. Harmelen, From SHIQ and RDF to OWL: The making of a web ontology language, *Journal of Web Semantics*, 2003.
[36] W. Hu, G. Cheng, D. Zheng, X. Zhong and Y. Qu, The results of falcon-AO in the OAEI 2006 Campaign, *Intl. Workshop on Ontology Matching (OM-2006)*, Athens, GA, USA, 2006.
[37] M. Klein, D. Fensel, F. V. Harmelen and I. Horrocks, The relation between Ontologies and XML schemas, *Linkoping Electronic Articles in Computer and Information Science* **6**(4) (2001).
[38] D. Kolas, J. Hebeler and M. Dean, Geospatial semantic web: Architecture of ontologies, *GeoS 2005*, LNCS 3799, 2005.
[39] L. V. S. Lakshmanan and F. Sadri, Interoperability on XML data, *International Semantic Web Conference '03*, 2003.
[40] Y. Li, J. Li, D. Zhang and J. Tang, Result of ontology alignment with rimom at oaei 2006, *Intl. Workshop on Ontology Matching (OM-2006)*, Athens, GA, USA, 2006.
[41] K. Lin and B. Ludäscher, A system for semantic integration of geologic maps via ontologies, *Semantic Web Technologies for Searching and Retrieving Scientific Data (SCISW)*, 2006.
[42] M. Lutz, Ontology-based descriptions for semantic discovery and composition of geoprocessing services, *GeoInformatica* **11**(1) (2007) 1–36.
[43] M. Lutz and E. Klien, Ontology-based retrieval of geographic information, *International Journal of GIS* **20**(3) (2005).
[44] J. Madhavan, P. A. Bernstein and E. Rahm, Generic schema matching with cupid, *Proc. 27th Intl. Conf. on Very Large Data Bases (VLDB)*, Rome, Italy, 2001, pp. 49–58.
[45] Z. Malik, A. Rezgui and A. K. Sinha, Ontologic integration of geoscience data on the semantic web, *International Geoinformatics Conference*, San Diego, CA, USA, 2007.
[46] C. Masolo, S. Borgo, A. Gangemi *et al.*, The Wonderweb Library of Foundational Ontologies, Preliminary Report, ISTC-CN.
[47] L. Mastella, Y. Ameur, M. Perrin and J.-F. Rainaud, Ontology-based model annotation of heterogeneous geological representations, *WEBIST* **(2)** (2008) 290–293.
[48] D. L. McGuinness, P. Fox, L. Cinquini *et al.*, Ontology-enabled virtual observatories: Semantic integration in practice, *Proceedings of the Fifth International Semantic Web Conference*, Athens, Georgia, November, 2006.
[49] B. Medjahed and A. Bouguettaya, A multilevel composability model for semantic web services, *IEEE Trans. Knowledge and Data Engineering* **17**(7) (2005) 54–68.
[50] P. Murray-Rust and H. S. Rzepa, Chemical markup, XML and the World Wide Web. 2. information objects and the CMLDOM, *Journal of Chemical Information and Computer Sciences* **41** (2001).
[51] U. Nambiar, B. Ludaescher, K. Lin and C. Baru, The GEON portal: Accelerating knowledge discovery in the geosciences, *Eighth ACM International Workshop on Web Information and Data Management* (*WIDM '06*), Arlington, Virginia, USA, November 10, 2006.
[52] I. Niles and A. Pease, Towards a standard upper ontology, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, C. Welty and B. Smith (eds.), Ogunquit, Maine, October 17–19, 2001.
[53] L. M. G. Olsen, S. Major, K. Leicester *et al.*, NASA/Global Change Master Directory (GCMD) Earth Science Keywords, Version 4.2.2.

[54] J. O'Sullivan, D. Edmond and A. T. Hofstede, What's in a service? Towards accurate description of non-functional service properties, *Distributed and Parallel Databases* **12**(2–3) (2002) 117–133.
[55] M. P. Papazoglou and D. Georgakopoulos, Serive-oriented computing, *Communcications of the ACM* **46**(10) (2003) 25–65.
[56] V. Parekh, J.-P. Gwo and T. Finin, Ontology based semantic metadata for geoscience data, *IKE 2004*, 2004, pp. 485–490.
[57] C. Phytila, An Analysis of the SUMO and Description in Unified Modeling Language.Phytila-SUMO.htm, 2002.
[58] L. Pouchard, A. Woolf and D. Bernholdt, Data grid discovery and semantic web technologies for the earth sciences, *Int. J. on Digital Libraries* **5**(2) (2005) 72–83.
[59] E. Rahm and P. A. Bernstein, A survey of approaches to automatic schema matching, *VLDB Journal* **10**(4) (2001) 334–350.
[60] D. W. Rankin, Appalachian salients and recesses: late precambrian continental breakup and the opening of the Iapetus Ocean, *Journal of Geophysical Research* **81**(32) (1976) 5605–5619.
[61] F. Reitsma and J. Albrecht, Modeling with the semantic web in the geosciences, *IEEE Intelligent Systems* **20**(2) (2005) 86–88, March/April 2005.
[62] T. Rodrigues, P. Rosa and J. Cardoso, Mapping XML to existing OWL ontologies, *International Conference WWW/Internet 2006*, pp. 72–77.
[63] E. Rusty, *XML in a Nutshell: A Desktop Quick Reference* (O'Reilly, 2002).
[64] A. T. Schreiber, B. Dubbeldam, J. Wielemaker and B. Wielinga, Ontology-based photo annotation, *IEEE Intelligent Systems* **16**(3) (2001) 66–74.
[65] L. Salayandia, P. Pinheiro da Silva, A. Gates and F. Salcedo, Workflow-driven ontologies: An earth sciences case study, *e-Science 2006*, **17**, 2006.
[66] S. Semy, M. Pulvermacher and L. Obrst, Towards the use of an upper ontology for U.S. Government and military domains: An evaluation, The Mitre Corporation (04-0603), 2004.
[67] A. Sheth, Changing focus on interoperability in information systems: From system, syntax, structure to semantics, in *Interoperating Geographic Information Systems*, 1998, pp. 5–30.
[68] A. Sheth, C. Bertram, D. Avant *et al.*, Managing semantic content for the web, *IEEE Internet Computing* **6**(4) (2002) 80–87.
[69] P. Shvaiko, A survey of schema-based matching approaches, *Journal on Data Semantics IV* (2005) 146–171.
[70] B. Simons, E. Boisvert, B. Brodaric *et al.*, in GeoSciML: Enabling the exchange of geological map data, In: AESC, Melbourne, 2006.
[71] A. K. Sinha, The latest Precambrian magmatic record of extensional tectonics in the central and southern Appalachians: Response to Grenville compressions, in Geological Society of America Abstracts with Programs, *Southeastern Section* **24**(2) (1992) 65–66.
[72] A. K. Sinha, J. Najdi, K. Lin *et al.*, Technical report01-2006, Element and Isotope ontologies for Geoscience, http://geon.geol.vt.edu/geon/pubreps/Technical report 01-2006.doc, 2006.
[73] A. K. Sinha, Z. Malik, A. Rezgui and A. Dalton, Developing the ontologic framework and tools for the discovery and integration of Earth science data, *Cyberinfrastructure Research at Virginia Tech (Annual Report)*, June, 2006.
[74] A. K. Sinha, Z. Malik, A. Rezgui *et al.*, Ontology Packages for GeoScience, Technical Report. http://geon.geol.vt.edu/geon/pubreps/Ontology Packages for Geoscience.doc, 2007.

[75] A. K. Sinha, A. Zendel, B. Brodaric *et al.*, Schema to ontology for igneous rocks, in A. K. Sinha, (ed.), *Geoinformatics*, Special Paper 397, Geological Society of America, 2006, pp. 169–182.
[76] R. P. Tollo and J. N. Aleinikoff, Petrology and U-Pb Geochronology of the Robertson River Igneous Suite, Blue Ridge Province, Virginia — Evidence for Multistage Magmatism Associated with an Early Episode of Laurentian Rifting, *American Journal of Science* **296**(1) (1996) 1045–1090.
[77] A. Uementsietsidis, M. Arenas and R. J. Miller, Mapping data in peer-to-peer systems: semantics and algorithmic issues, *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, 2003, pp. 325–336.
[78] WSMO Working Group, Web Service Modeling Ontology (WSMO). http://www.wsmo.org/, 2004.